

## Like Powered Census

Dušan Barok

talk at Liwoli festival, Linz

13 May 2011

Recently we have seen the arrival of distributed/federated social platforms like *Diaspora*, *StatusNet*, *GNUSocial*, or *Thimbl*. These platforms are understood as the *critique* of the social network services. They are based on the premises that the *open source software* built upon *distributed database* and with *more advanced privacy controls* will provide the non-exploitative alternative to facebook, twitters and tumblrs.

I would like to question the *validity of this assumption*.

In order to do this, let us look at what may be the core of competition between social web oligarchs, Facebook and Google.

I've spent last months researching these issues at my studies at Piet Zwart Institute. My interest started with finding out about the small and marginal *conflict* between these two companies.

Last year in November, Facebook turned down Google's request for automatic import of Facebook friend list to Gmail. Google responded by blocking Facebook from one-click import of Gmail contacts.

What I realised was that they not only blocked the feature of exploiting one network to promote the other, but that they didn't want one network *access* the user's friend/contact lists at the other. They were *protecting* the friend lists.

To understand why the friend list is such a valuable thing for them, I had to learn more about how it is connected to the value extraction. It had to have something to do with their main source of income – *advertising*, obviously.

In 2009, 97% of Google's revenue came from advertising, and more than 80% in case of Facebook. Advertisers pay for their products placements on the websites plugged into and depending on these centralised resources. Whether it is within facebook.com and google.com domains, or across the web, relevance of ads is calculated from particular user's data. In last decade, these data consisted of profile information, such as location, age, gender, languages spoken, education, workplace, relationship status, and interests, such as likes, products, or searched terms. Personalisation of online services allowed the direct marketing to bloom. Users keep on producing their online identities, these are in turn being stored in centralised databases and rented to advertisers. This won't probably change anytime soon. Advertising proved to be the major income for pre-internet mass media too, only that with personalisation it got more sophisticated. And now with *socialisation* it is getting even more. Owning the information about how users are related to each other, I was sure that it was just a matter of time when the network providers come up with solutions to monetize the *social relations* too.

One of the attempts to do this was Facebook Beacon launched in 2007. It allowed Facebook to collect data about user's purchases from external partner websites and make users share information about their purchases with their friends. Technically, the service used cookies on the client side and web bugs embedded within the third-party sites. It worked without consent of the users, what obviously raised

privacy concerns. The class action lawsuit followed and the service shut down two years later. It did not take too long to find the roundabout. After the spread of Like buttons across the web, earlier this year Facebook launched 'Sponsored Stories Ad Unit', one of the 'social ad' programs using friends to advertise the products. As the website explains: "Social adverts pair an advertiser's message with social actions you have taken, such as liking a Page. You only appear in social adverts to your confirmed friends."

To have the actual friends advertise products to the users sounds like a clever move. Then I learnt about concept of 'homophily' used in social theory, which confirms this. It basically says that the "people who communicate with each other are *more likely* to be similar to each other". Building upon this concept, Hill, Provost and Volinsky in their paper found that " 'network neighbours'—those consumers linked to a prior consumer—adopt the service [or product] at a rate 3-5 times greater than baseline groups selected by the best practices of the firm's marketing team". *Users are more likely to consume products previously consumed by their friends*. It is the ground assumption of the coming next wave of direct marketing: what they call '*network-based marketing*', or what Wired named '*social commerce*'. Authors of paper list examples of companies collecting the social data:

"eBay purchased Skype; they now have large-scale, explicit data on who talks to whom. With Gmail, Google has access to explicit networks of consumer interrelationships and is using Gmail for marketing; directed network-based marketing might be a next step. Various systems have emerged that provide explicit linkages between acquaintances (e.g., MySpace, Facebook). As more consumers create interlinked blogs, another data source arises. More generally, these results suggests that such linkage data potentially could be a sort of data considered for acquisition by many types of firms, as purchase data now are being collected routinely by many types of retail firms through loyalty cards." They close the paper with stressing the importance of social data over other personal data for marketers: "It may well be that *direct communications between people is a better indicator of deep similarity than any demographic or geographic attributes*."

This is probably it, this is why Facebook tends to protect their friends lists. It's their competitive advantage. How is it with *Google*?

Google was not successful with social applications: operation of social networking platform Orkut was fully transferred to Google Brazil, OpenSocial framework did not receive expected attention, development of Google Wave was suspended, and Google Buzz sharing tool was largely ignored for bringing nothing new and original. Largest revenue stream still comes from its search engine. To exploit social data from the users in order to feed the social ads, Google introduced Social Search in 2009. Users are encouraged to enter their usernames they have at various social network sites into their Google Profile accounts, so that the engine can give a special ranking to the relevant content generated by their friends and contacts across the social platforms, and favor it in search results. Basically, users get their search results annotated with social information in exchange for donating their social connections from across the web to Google.

This is only a part of network-based marketing story. Facebook and Google are also competing in serving as the supplier of population's social data to industries, which are being developed around them. To expand the sources for advertising revenues by including the social games and applications. They needed to develop methods of delivery, while preserving the control over access to these social data. They came up with the concept of *social graph*, which is basically the sum of the friend/contact lists of all users. Firstly, in May 2007, Facebook enabled developers to "build full social applications on top of the social graph, inside of Facebook."

A couple of months later, internet entrepreneur and LiveJournal creator Brad Fitzpatrick published a proposal to make “the *social graph a community asset*, utilizing the data from all the different sites, but not depending on any company or organization as ‘the’ central graph owner”. Fitzpatrick was subsequently hired by Google, which proudly launched its Social Graph *API* in January 2008, giving graph to the public. His plan was realised, although two fundamental conditions were omitted: API has *conditioned* access and data are collected and hosted *exclusively* on Google’s servers. Here is what he called for in original proposal:

“a. Establish a *non-profit and open source software* (with copyrights held by the non-profit) which collects, merges, and redistributes the graphs from all other social network sites into one global aggregated graph. This is then made available to other sites (or users) via both public APIs (for small/casual users) and downloadable data dumps, with an update stream / APIs, to get iterative updates to the graph (for larger users)

b. While the non-profit’s servers and databases will initially be centralized, ensure that the design is such that *others can run their own instances*, sharing data with each other. Think ‘git’, not ‘svn’. Then whose APIs/servers you use is up to you, as a site owner. Or run your own instance.” (Fitzpatrick 2007)

Facebook responded in April 2010 by introducing Graph API, containing much larger social graph, and expanding it by objects including photos, events, and product pages. While both APIs employ the open source standards for data structure and authorisation (XFN, FOAF, JSON, OAuth), the data sets are stored centrally and access to them is limited. Facebook allows third parties to access user data only upon the temporary approval by the user (unless user defaulted them to public), and Google provides only “publicly declared connections”.

I also realised that monopolisation of social graph goes hand in hand with privacy control. And can say that social platforms willingly operate in the linguistic greyzone between the *public and private*. Their business model depends on providing customers data which are unique and which they don’t find elsewhere. There was another interesting event. The Facebook’s model of data enclosure was challenged by exponential rise of Twitter’s user base, which embraces the model of controlled open space. Consequently, in December 2009 it announced opening up all profiles with photos and other social data to whole internet, even at a price of reinventing their revenue model. Serving as a role model, Zuckerberg opened up his own profile “sharing photos of himself at parties and with his girlfriend”. To his surprise, the user base went on strike to demand “privacy” and pushed Facebook back to readapt their business model of secrecy and closed data, thus legitimising it.

Google follows the example. In an update to Social Search they “given you more control over how you connect accounts, and made connecting accounts more convenient. You can still connect accounts publicly on your Google profile, but now we’ve added a new option to connect accounts privately in your Google Account. (After all, you may not want everyone to know you’re @spongebobsuperfan on Twitter)”. Playing on the privacy tune, Google gives its users protection of their social profile from others.

So thus I learnt that by being offered the privacy control settings within the network, the users *perform their privacy* and voluntarily feed in the content designated solely for their peers. This creates not only “walled gardens” of closed systems, but more importantly, *privacy lock-in* for users who are left to demand protection of their personal data.

*As a result, by creating a problematic private/public divide, the network owners are justified to take upon the role of protector, "privatise" the private data and enclose the social graph generated in this way.*

Coming back to my original question, I was curious whether the combination of open source software, distributed database and advanced privacy controls is enough to provide the non-exploitative alternative to what I have just described.

My question is how should we treat the social graph. Should we treat it as a public good as Fitzpatrick originally proposed? And if it is a public good, does the data about who we talk to and who we friend have other use rather than for marketing and statistical analysis?

I'm asking this because there *are* tools for online communication and collaboration which do not know such a thing as friends list. We can think of *mailing lists*, *irc*, or *wikis*. Where people are part of the group by subscribing to it, and talk and work together without the requirement of declaring the friend or acquaintance connections to particular others.

I will not finish with the grand conclusion, only with a question I'm puzzled with right now. Shall the federated social platforms promote the open social graph, in other words social graph as a community asset, or, shall we rather focus on the critique of the very social graph and view it as the commodified social relations.